



The Electronic Text Center at the University of Virginia Library

著者	Seaman David, Stubbs Kendon
雑誌名	デジタル図書館
号	9
ページ	55-62
発行年	1997-03-05
URL	http://hdl.handle.net/2241/102936

The Electronic Text Center at the University of Virginia Library

David Seaman
Director, Electronic Text Center
Alderman Library
University of Virginia
Charlottesville, VA 22903, U.S.A.
Tel.: 804-924-3230 Fax: 804-924-1431
E-mail: etext@virginia.edu
URL: <http://etext.lib.virginia.edu>

Kendon Stubbs
Associate University Librarian
Alderman Library
University of Virginia
Charlottesville, VA 22903, U.S.A.
804-924-3026 Fax: 804-924-1431
E-mail: kstubbs@virginia.edu

Abstract

The Electronic Text Center is a library-based humanities computing service, with two main areas of activity: it builds and maintains an on-line archive of electronic texts and images, and it builds and maintains a broad-based user community in the humanities. After five years of activity, we have a large body of online SGML texts in a variety of languages, and after hundreds of training sessions we have a body of student and faculty users who are making sophisticated use of our services in their teaching and research.

Keywords

Electronic Text Center : electronic text : SGML : HTML : digital image : full-text database : digital library : on-line library : humanities computing : internet training : teaching and technology.

INTRODUCTION

After nearly five years of operation, the Electronic Text Center continues both to develop its on-line collections and to encourage the use of electronic texts at the University of Virginia and beyond. The Electronic Text Center was created by the Library to deal with the first wave of commercial humanities-related electronic texts that were published in the early 1990s, including the *Oxford English Dictionary*, the *English Poetry Database*, and the *Patrologia Latina*. In 1992, the University of Virginia Library committed space, equipment, and staff to create a Center as a means to encourage and develop humanities computing, long before there was any pressure from our users to provide such a service. The Center makes available equipment for the creation and analysis of electronic text; it provides training for these new tools and techniques; it acts as a focal point for SGML development in the humanities at Virginia; and it provides a place in which to use those texts that are not yet accessible on the Internet.

ELECTRONIC TEXT HOLDINGS

The Internet-accessible holdings now contain many thousands of texts and related digital images. Some, such as the following, are commercially-available items, held by us under contracts that limit their use to the University of Virginia or to the 39 VIVA (Virtual Library of Virginia) sites. We have always bought these large commercially-available electronic text collections as SGML data files rather than as CD-ROMs, in order to load them onto a Unix server and provide on-line access to our users across the Internet:

The Oxford English Dictionary

(25 volumes)

English Poetry Database

(4,500 works)

English Verse Drama Database

(1,500 titles)

English Prose Drama Database

(1,500 titles)

American Poetry Database

(4,500 works)

African-American Poetry Database

(2,500 poems)

Others, such as the following, are limited in their use to University of Virginia:

The Patrologia Latina

(221 volumes)

The Old English Corpus

(3,000 items)

British philosophy: 1600-1900

American Civil War newspapers

In addition, there are many hundreds of other literary, historical, philosophical, and religious materials in a variety of languages. The selection includes many 18th- and 19th-century English literary and historical works (often with illustrations) [slide I], and some French, German, Latin, Japanese, and medieval English titles. Most of this latter group of texts can be accessed by any Internet user; currently we receive about a million hits a month on our web server, from all over the world. To put that in daily terms, we receive about 33,000 hits on about 15,000 pages from about 4,000 different host machines.

All of our on-line texts are encoded with Standard Generalized Markup Language (SGML) and are converted automatically to HTML for use through the World Wide Web. Those texts we create or mark up ourselves are tagged according to the Text Encoding Initiative Guidelines by the staff of the Center, and we work with increasing frequency in partnership with students and faculty at the University of Virginia and at other universities to create new electronic texts and images. A good example of a collaboration to produce new texts is the *Japanese Text Initiative* [slide II], a growing collection of SGML-encoded and searchable Japanese literature published on the Internet by the University of Virginia Library Electronic Text Center and prepared by Kendon Stubbs at the University of Virginia and Sachie Noguchi at the University of Pittsburgh.

Among the English language holdings are texts and images taken from our Special Collections, including numerous Jefferson letters, Mark Twain material, and 19th-century African-American historical documents [slide III]. As faculty and students use the Electronic Text Center itself they are moving from being consumers of online information to become producers of it, re-shaping and supplementing our holdings into clusters of web documents that suit a particular purpose. Professor Stephen Railton's *Mark Twain in his Times*, a teaching tool for the study of the American 19th-century novelist, is a good example of how the electronic medium is being used by our faculty [slide IV].

TEXTS AVAILABLE OFF-LINE

While the majority of our holdings are available through the World Wide Web, there are some that we cannot network for legal or technical reasons, including the *Global Jewish Database*; the *Thesaurus Linguae Graecae* (8,000 works of ancient Greek); *Perseus* (a hypertext collection of Greek texts and images); *CETEDOC* (Latin theological works); *Admyte* (medieval Spanish); the works of Robert Musil, of Immanuel Kant, and of Thomas Aquinas; and selected 19th-century American Poetry.

ACCESS

Because most of our electronic texts are available on-line, we can provide the same search software (from OpenText Corporation) for all our collections, and we can use a Web browser such as Netscape as a common and familiar interface. We have developed our own suite of CGI scripts, including web forms and SGML-to-HTML filters, to allow our users to search and browse the texts. The benefits to the user are obvious: having been taught to use one database, a user knows how to search any of our on-line holdings, thereby overcoming the frustrations involved with using CD-ROMs, where each disk has a different interface.

USERS

A principal aim of the Center is to build a broad-based user community for humanities-related electronic resources at Virginia. To this end we run regular training sessions, including classes on scanning, HTML, and on other aspects of the use and creation of electronic texts and images. For five years we have worked daily with individual users who range from first-year undergraduates in composition classes to graduate students studying aspects of Anglo-Saxon literature, American Studies, rabbinical responsa, medieval French, and various other teaching and research projects. We are also pleased to assist the Fellows of UVa's Institute for Advanced Technology in the Humanities.

Increasingly, we are building relationships with university presses, academic publishers, the producers of scholarly journals, and other emerging digital libraries. Our server provides support and online publication space for two journals edited and produced at the University of Virginia – *The Visual Anthropology Review and Essays in History* – and we have produced searchable etexts of two University Press of Virginia titles: Timothy D. Pyatt's *Guide to African-American Documentary Resources in North Carolina*. (Charlottesville, 1996) and Michael Plunkett's *Afro-American Sources in Virginia. A Guide to Manuscripts* (Charlottesville, 1995) [slide V]. Currently, we are finishing the conversion of 50 years of the journal *Studies in Bibliography* to SGML files, for distribution on the Internet, as part of an ambitious electronic publication program underway by the Bibliographical Society of the University of Virginia.

Moreover, we have been successful in attracting grant money to help underwrite the cost of two ambitious projects: *The American Heritage Virtual Archive Project (EAD)* and *The Electronic Archive of Early American Fiction*. The latter is the subject of another paper at this conference; the former is a national Endowment for the Humanities (NEH) grant to Virginia, Duke, Stanford and UC Berkeley to encode thousands of pages of archival finding aids in SGML, using the EAD – Encoded Archival Descriptions – tags.

Increasingly, the Electronic Text Center provides a model for other institutions as they plan similar endeavors. Scores of librarians and scholars have visited the Center, including groups from Harvard, Duke, Indiana, Johns Hopkins, Iowa, Yale, Columbia, Chicago, Kentucky, UC Berkeley, Virginia Tech, Richmond, UNC Chapel Hill, UT Austin, Emory, the National Humanities Center, the British Library, Oxford, Cambridge, Nottingham, Glasgow, Leiden, Bielefeld, Groningen, the National Diet Library, Osaka University of Foreign Studies, and ULIS in Japan, and Sydney, Macquarie, and Curtin universities, Australia. This activity is important to us, as it fosters the development of electronic text and image services elsewhere.

In our next five years we expect to see the same tremendous pace of change that has marked our first half-decade as a library service, with emerging technologies such as Virtual Reality Modeling Language (VRML) and Internet-wide search tools becoming increasingly useful to the humanities scholar. Our user community is getting increasingly sophisticated in their ambitions and needs, and the amount of full-text and image data coming out from commercial publishers is growing dramatically. The fundamental notions on which our service was founded continue to hold true throughout all this growth: the data must be held in standardized forms such as SGML that can survive in a time of rapid technological change, and the library's role and skills are central to the development and long-term viability of on-line services.

WORKS CITED

The Modern English Collection

<http://etext.lib.virginia.edu/modeng/modeng0.browse.html>

British Poetry 1780-1910

<http://etext.lib.virginia.edu/britpo.html>

Mark Twain in his Times

<http://etext.virginia.edu/railton/>

Japanese Text Initiative

<http://etext.lib.virginia.edu/japanese/>

Electronic texts from Special Collections

<http://etext.lib.virginia.edu/speccol.html>

Timothy D. Pyatt. *Guide to African-American Documentary Resources in North Carolina*. Charlottesville, 1996.

<http://www.upress.virginia.edu/epub/pyatt/>

Michael Plunkett. *Afro-American Sources in Virginia. A Guide to Manuscripts*. Charlottesville, 1995.

<http://www.upress.virginia.edu/plunkett/>

Essays in History

<http://etext.lib.virginia.edu/journals/eh/>

The Visual Anthropology Review

<http://etext.lib.virginia.edu/VAR/>

SGML Resources

<http://etext.lib.virginia.edu/sgml.html>

VIVA, the Virtual Library of Virginia

<http://www.viva.lib.va.us/>

The Bibliographical Society of the University of Virginia

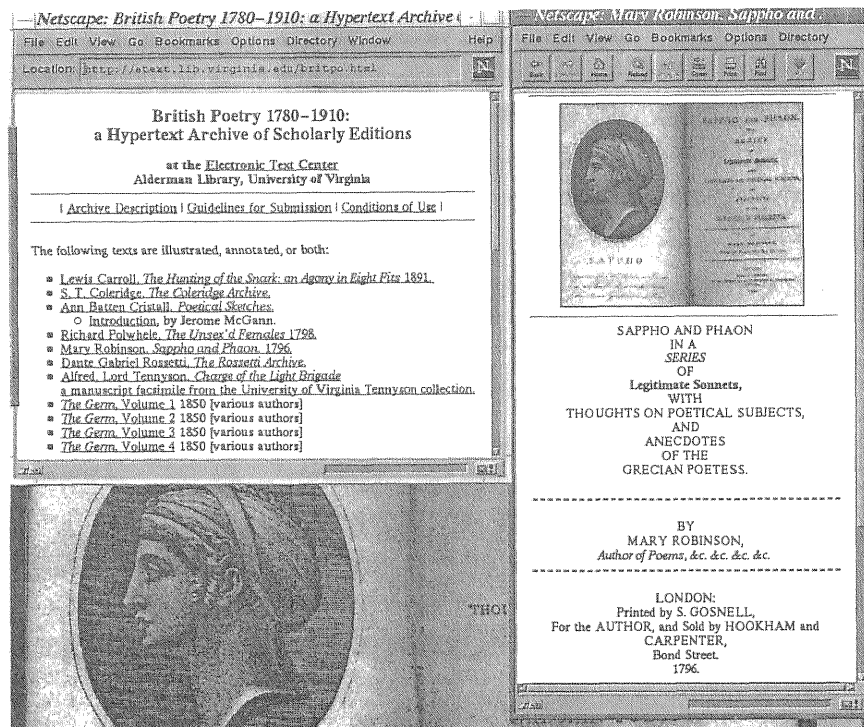
<http://etext.lib.virginia.edu/bsuva/>

The Electronic Archive of Early American Fiction

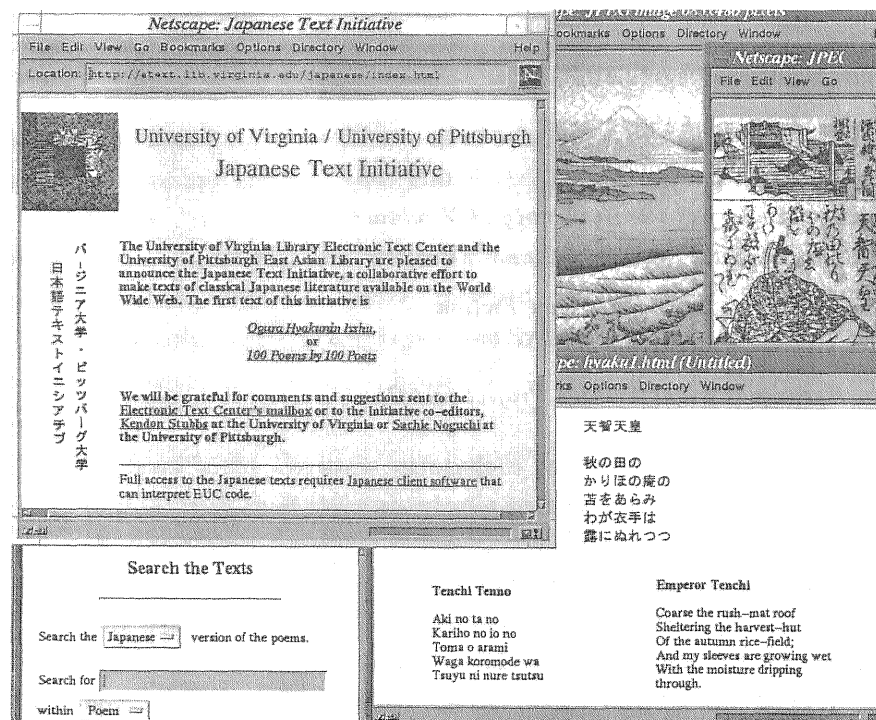
<http://etext.lib.virginia.edu/eaf/>

The American Heritage Virtual Archive Project (EAD)

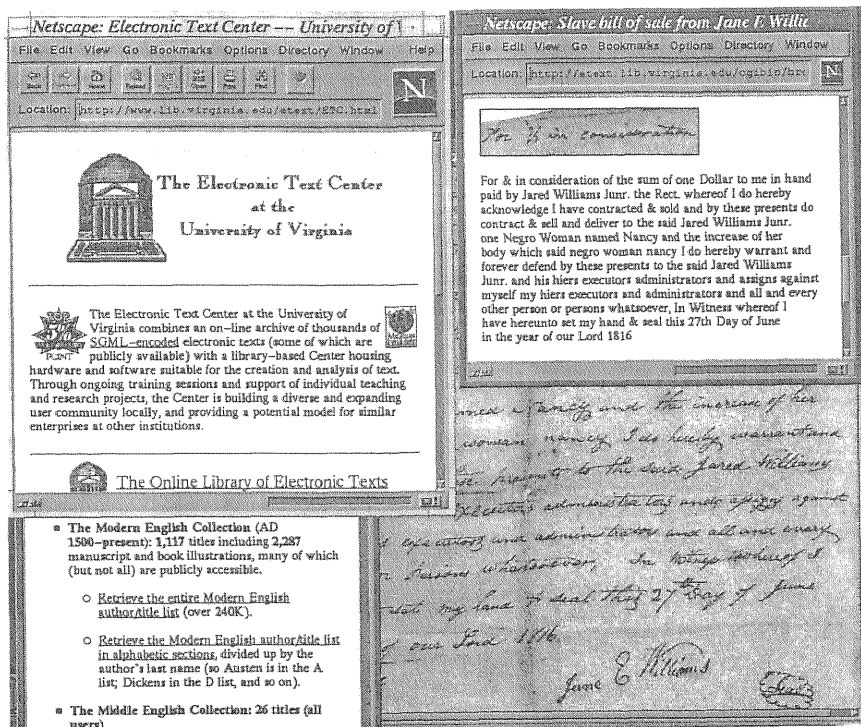
<http://etext.lib.virginia.edu/ead/>



Slide I



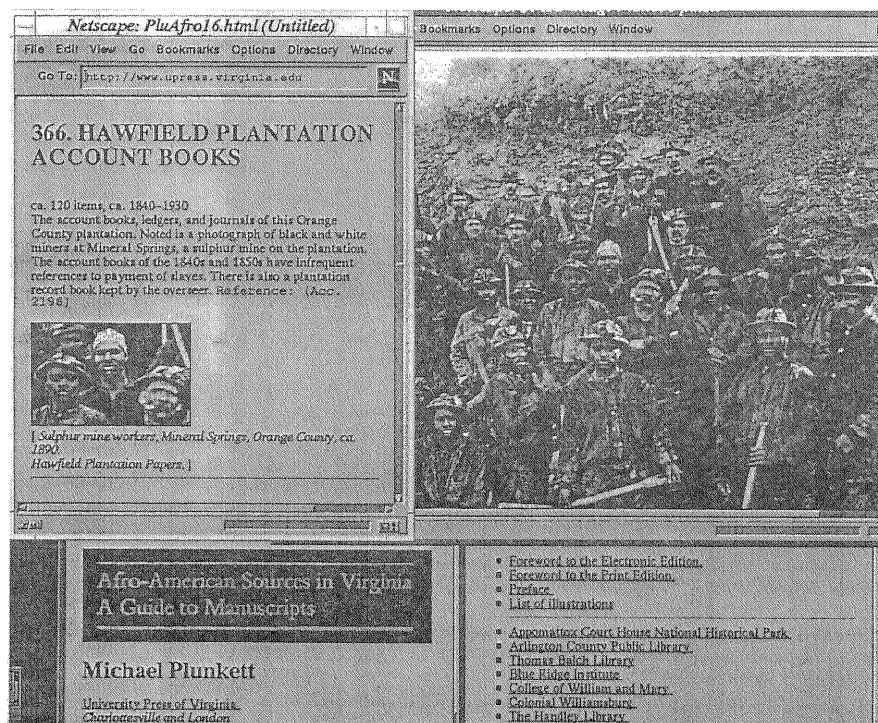
Slide II



Slide III



Slide IV



Slide V